Real Time Speaker Identification System – Design, Implementation and Validation

Dr. M. Meenakshi
Professor, Dept. of Instrumentation Technology, Dr. AIT Bangalore 560056
Email: meenakshi_mbhat@yahoo.com

Abstract— **This paper presents design, implementation and validation of a PC based Prototype speaker recognition and verification system. This system is organized to receive speech signal, find the features of speech signal, and recognize and verify a person using voice as the biometric. The system is implemented to capture the speech signal from microphone and to compare it with the stored data base using filter-bank based closed-set speaker verification system. At first, the identification of the voice signals is done using an algorithm developed in MATLAB. Next, a PC based prototype system is developed and is validated in real time. Several tests were made on different sets of voice signals, and measured the performance and the speed of the proposed system in real environment. The result confirmed the use of proposed system for various real time applications.**

Keywords— **Biometrics, voice recognition, feature extraction, finger print, MATLAB, auto correlation, Euclidean distance.**

I INTRODUCTION

Speech recognition has been an active field of research during the last three decades. Speaker Recognition is a process of automatically recognizing who is speaking on the basis of speaker dependent features of the speech signal. Basically, speaker recognition is classified in to speaker identification and speaker verification. Wide application of speaker recognition system includes control access to services such as banking by telephone, database access services, voice dialing telephone shopping so on. Now, speaker recognition technology is the most suitable technology to create new services that will make our every day lives more secured.

Biometrics is seen by many researchers as a solution to a lot of user identification and security problems [1]. This may include speaker identification, face recognition, fingerprint recognition, finger geometry, hand geometry, iris recognition, vein recognition, and voice and signature recognition. Various techniques are available in literature to resolve the automatic speaker identification problem [2 - 3]. Generally, speech signal is mixed up with noise signal. However, most of the work on speech processing for the speaker recognition was done by focusing the speech under the noiseless environments and only few by focusing speech under noisy conditions [4-5].

Researchers in [6 -8] are used algorithms based on Cepstral analysis or homomorphic analysis for automated speech recognition. Most frequently used method is the statistical hidden Markov model (HMM) as in [9-10] which use libraries of words and grammar rules to select the highest probability outcome from a sequence of samples. The cepstral analysis supplanted the direct use of linear prediction analysis LP, derived from the hidden Markov modeling.

A speaker recognition system often works in either of two operating modes i.e., Text-dependent, where the same or known text is used for training and test and text-independent. This paper covers the detailed study of MATLAB simulation of the filter-bank based closed-set speaker verification system. It involves the study of filters namely the FIR and IIR filter design and the study of efficiency of the filter-bank based speaker verification system. Here, algorithm for “text-dependent Speaker Verification”, which may be employed in security systems, is developed. Finally a real time PC based prototype system for automatic opening/closing of the door is developed and validated.

This research is carried out under two headings i.e. training and recognition. In the training phase” fingerprints” of the voice signal are extracted and stored in database. “Fingerprint”, which is a vector of numbers, represents characteristics of the sound in the frequency domain as time evolves. Each number represents the energy, or average power that was heard in a particular frequency band, during a particular interval of time. In the recognition phase, the fingerprint of the immediate voice input is compared with finger-prints stored in the database. Here the efficiency and drawbacks of two comparison processes namely: Euclidean distance and Auto Correlation Co-efficient methods are examined.

The organization of this paper is as follows: Section II highlights the principles of the proposed speaker identification system. Implementation methodology of the proposed system is given in section III. Next, section IV highlights the results and analysis of the developed system. Real time validation of the PC based, speaker identification system is also presented in section IV. Finally conclusions are drawn in section V.

II PRINCIPLES OF THE PROPOSED SPEAKER IDENTIFICATION SYSTEM

Voice recognition refers to the process of identification of a person’s identity with voice as the biometric. The objective of speaker identification system is labeling an unknown voice as one of a set of known voices. Architecture of the speaker identification system depends on the choice of recognition situation. Probabilistic approach is most suitable in the case of text independent situations whereas a

DOI: 01.IJCSI.02.01.82

time alignment, the dynamic time warping (DTW) of the utterance with the test can be enough for text dependent cases.

The focus of this work ison the identification of a speaker from a group of N known speakers, in a text independent situation. Speaker Verification is the process of determining whether the speaker identity is who the person claims to be. One among the various approaches to speech recognition is the pattern recognition approach, which has two steps, namely, training of speech patterns, and recognition of patterns via pattern comparison. The concept is that if enough versions of a pattern to be recognized are included in the training set provided to the algorithm, the training procedure should be able to adequately characterize the acoustic properties of the pattern. This type of characterization of speech via training is called as pattern classification. Here the machine learns, which acoustic properties of the speech class are reliable and repeatable across all training tokens of the pattern. The utility of this method is the pattern comparison stage with each possible pattern learned in the training phase and classifying the speech according to the accuracy of the match of the patterns. The basic structure of speaker verification system, which uses the above said approach is shown in Fig. 1 and has three main components, i.e, Front-end processing, Speaker modeling and Pattern Matching.

A. Front-end processing: This refers to training phase and is used to highlight the relevant features and remove the irrelevant ones. This phase is accomplished using the filter-bank based method of implementation and tested with both FIR and IIR filters.

B. Speaker Modeling: This phase mainly involves storing of relevant features extracted in the Front-end processing step.

C. Pattern Matching: This basically involves matching or comparison of the stored speaker model with the input speaker model (voice from unknown speaker) .Here, Euclidean Distance and Auto Correlation Coefficient methods are used for the comparison.

III. METODHOLOGY

The MATLAB simulation of the speaker verification system based on filter-bank method involves the following steps:

1. Training Phase that includes: (A) Speech Spectrum Analysis, (B) Decimation and FFT, Band-Pass filtering and (C) fingerprint accumulation

2. Recognition Phase including: (A) Comparison and (B) Decision making

The speech input is read into MATLAB at sampling frequency of 8 KHz since human speech spectrum lies under 4 KHz. Signal, which is read in at 8000samples/second is reduced to 4000samples/second by decimation of input signal by a factor of 2. The decimated signal is then interpolated by the same factor to validate the decimation process as in Figure. 2. Once the decimation is completed, the FFT of signal is obtained to get the frequency domain values of the speech signal.

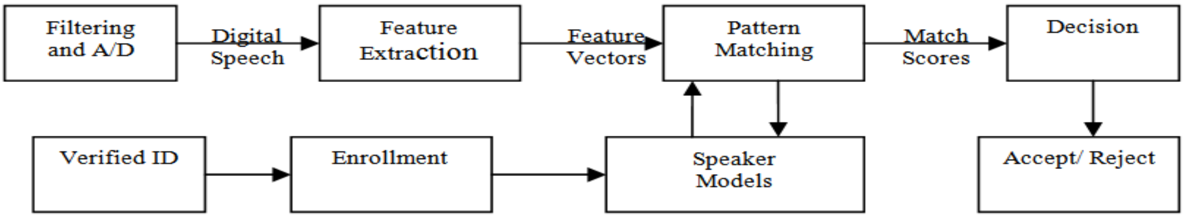

Figure 1. Basic Structure of Speaker Verification System

In the next step, i.e., bands pass filtering and finger print accumulation, the method employed is termed as “sub-band filter bank” coding, in which the speech signal is first split into frequency bands using a bank of band-pass filters.

The filter bank is a collection of band-pass filters all processing the same input signal. Here, three different frequency bands are selected namely: 50-500Hz, 500-1000Hz and 1000-1500Hz. Also, the band pass filtering is tested with the IIR and FIR filters and the performance of these two implementations is gauged. Figs. 3 and 4 gives the outputs of FIR and IIR filters for the frequency of 0-500Hz with, the sampling frequency of 2 kHz.

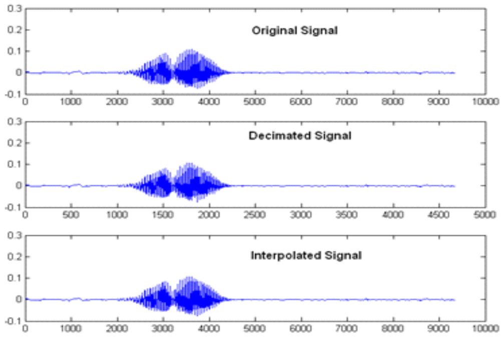

Figure 2. Decimation and interpolation of speech signal

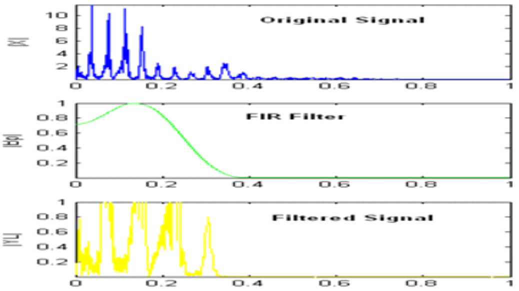

Figure 3. FIR filtering of speech signal

DOI: 01.IJCSI.02.01.82

Next, the output of the bandpass filters is accumulated to form the finger print of the voice signal, which is obtained by the summation of the square of the bandpass filter outputs.

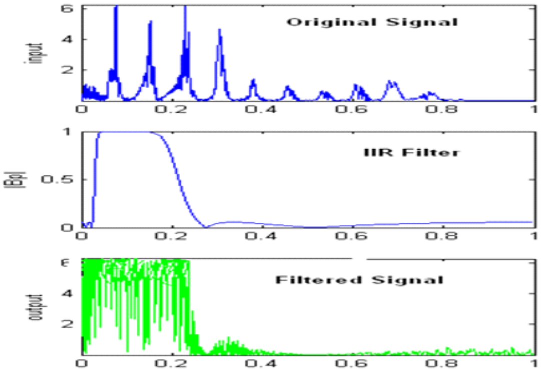

Figure 4. IIR filtering of speech signal

The next step is comparison and is done by using Euclidean distance method and Auto Correlation Coefficient method. Minimum Euclidean distance indicates a closer match between the two signals compared. Similarly, auto Correlation of +1 indicates maximum match and a –ve value or a value close to 0.0 indicates no match.

IV. RESULTS AND ANALYSIS

Figs 5 and 6 give the graphical representation of the simulated results of the comparison using Euclidean distance and Auto correlation coefficient methods respectively.

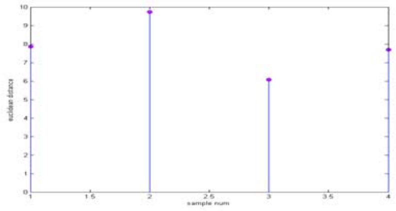

Figure . 6. Euclidean distance of same word spoken by 4 different users

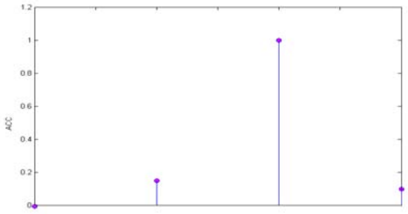

Figure 7. Auto Correlation Coefficient of same word spoken by 4 different users

Fig 6 indicates minimum Euclidean distance to person 3 hence, selected user is person 3. Also, note that the Euclidean distance of other users are very close to that of person 3. Hence, when there are slight variations in the users voice input, the Euclidean distance changes due to which the system fails to recognize the correct user. Also, when the user's voice is imitated by others, the comparison gives wrong identification. Therefore the accuracy achievable using Euclidean distance is reduced to 65-70%. Hence, it is concluded that Euclidean distance method of comparison is not the best ways for comparison.

From the Fig. 4, it is clear that auto correlation coefficient for person 3 is 1 and the other users are very low. Hence user 3 is selected by the system. Results demonstrate that the probability of matching with a mimic signal was considerably reduced. It gave an accuracy of more than 85%. The recognition accuracy was high even if there were slight variations. Thus, it proves to be a better method of recognitions than Euclidean distance method.

Finally, several real time voice signals are inputted to the PC Based Speaker Identification/ Recognition System and validated the above mentioned results. A prototype speaker identification system is developed to open or close the door automatically by receiving the known voice signal and validated in real time.

IV. CONCLUSION

This paper explained the development of simulated speaker verification system for the speaker identification application. Recognition is performed by Filter-bank based method. Filter banks are designed to reduce the noise and to extract essential features of speech signal. Higher the number of bands selected, higher the accuracy obtained. Finally the proposed algorithm is validated in real time PC based, prototype speaker identification system. It is proved that, the proposed system is suitable for different real time applications.

References

[1] A. Jain, R. Bole, S. Pankanti, *"Biometrics Personal Identification in Networked Society"*, Kluwer Academic Press, Boston, 1999.

[2] Jain, A., R.P.W.Duin, and J.Mao., *"Statistical pattern recognition: a review"*, IEEE Trans. on Pattern Analysis and Machine Intelligence 22, pp. 4–37, 2000

[3] Sadaoki Furui, *"50 Years of Progress in Speech and Speaker Recognition Research"*, ECTI transactions on computer and information technology, Vol.1, No.2, 2005.

[4] Zoubir Hamici, *"Speaker Recognition Using Spectral Cross-correlation: A Fast Algorithm"*, Journal of Computer Science (Special Issue): pp. 84-88, 2005

[5]Wu, D., Morris, A.C.&Koreman, J.,*"MLP Internal Representation as Disciminant Features for Improved Speaker Recognition"*, in Proc. NOLISP2005, Barcelona, Spain, pp. 25-33, 2005.

[6] A. Mesaros, and J. Astola, *"The mel-frequency cepstral coefficients in the context of singer identification,"* in Proc. of ISMIR *2005*, London, UK, pp.11-15, September 2005.

[7] Donato Impedovo, Mario Refice, *"Optimizing Features Extraction Parameters for Speaker Verification"* 12th WSEAS International Conference on SYSTEMS, Heraklion, Greece, pp 498 -503, July 22-24, 2008

[8] D. Impedovo, M. Refice, *"The Influence of Frame Length on Speaker Identification Performance"*, Proceedings of IAS 2007, Manchester, 2007.

[9] C.Y. Espy-Wilson, S.Manocha, S.Vishnubhotla,, *"A new set of features for text-independent, speaker identification"*, Proceedings of ICSLP, 2006, pp.1475-1478, 2006.

[10]. Povlow, B. and S. Dunn,. *"Texture classification using noncausal hidden markov models"*, IEEE Trans. Pattern Analysis and Machine Intelligence, 17: pp. 1010 -1014., 1995.

DOI: 01.IJCSI.02.01.82